

DPM at ASGC

Shu-Ting Liao

ASGC

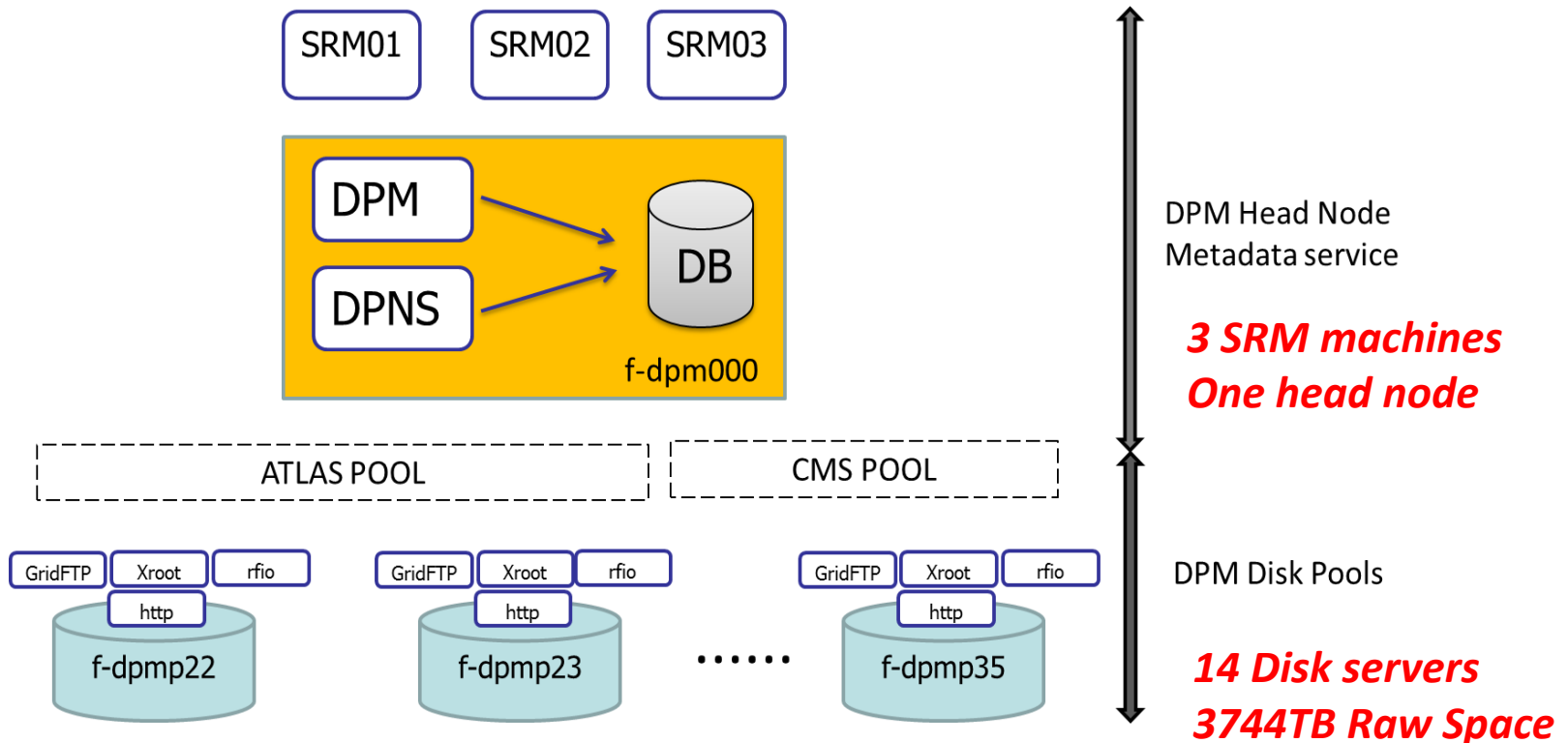
Edinburgh DPM Workshop 2013

Recent Updates

- Upgraded DPM to 1.8.7 on Nov 25
 - Support Webdav renaming for Rucio
 - New FAX n2n plugin for better performance
- Increased international network link to Amsterdam to 20Gb.
- Running Asia region redirector(atlas-xrd-asia.grid.sinica.edu.tw)

Current Status

- Mainly support ATLAS Tier1 Tier2 and CMS Tier2



Current Status

- Running another DPM instance for AMS experiment.
- AMS transfer compared to ATLAS:

- File Transfer Service via GRID
- SRM endpoints at CERN and TW
- Duration: Sep. 2012 – Aug. 2013
- CERN to TW

Total Volume: 1427.738TB

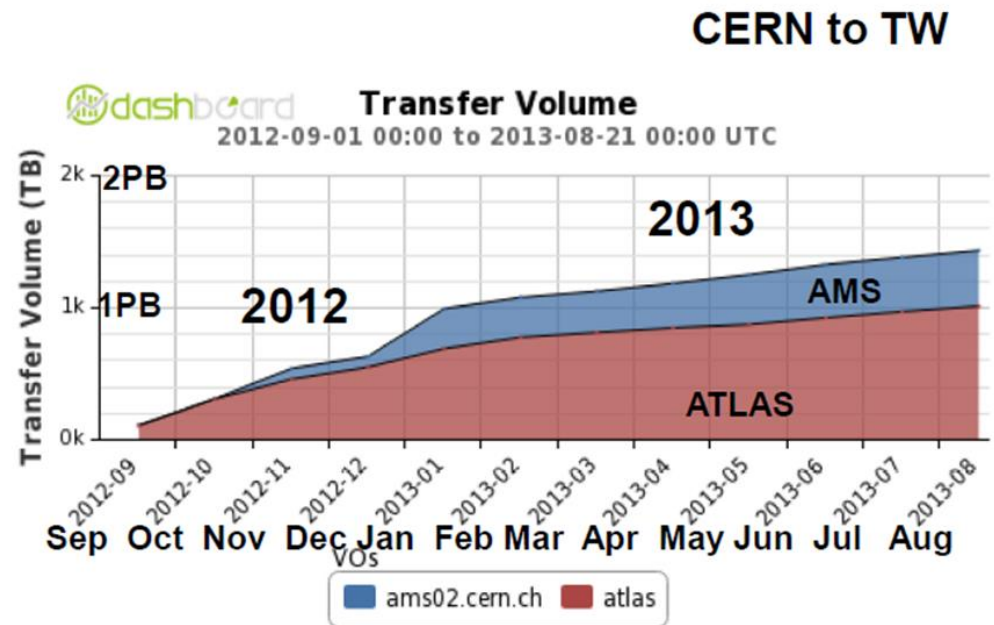
AMS: 423.654TB

ATLAS: 1,004.085TB

Total Number of File: 3,201,572

AMS: 598,815

ATLAS: 2,602,757



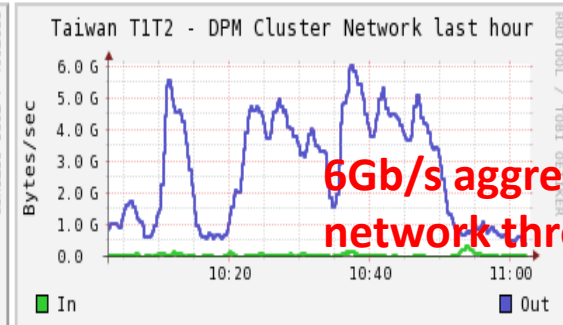
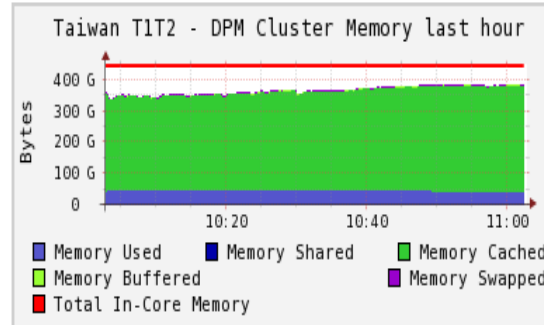
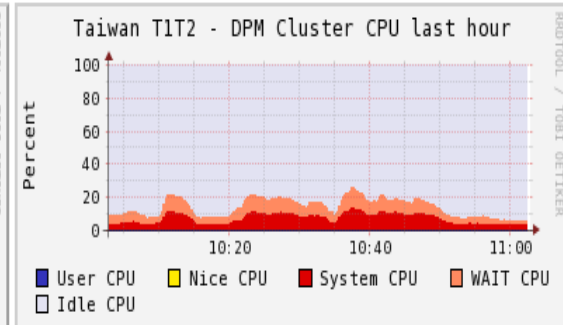
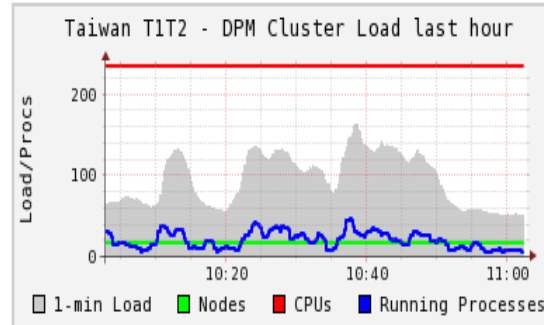
Performance (I)

Overview of Taiwan T1T2 - DPM

CPU's Total: 236
 Hosts up: 18
 Hosts down: 0

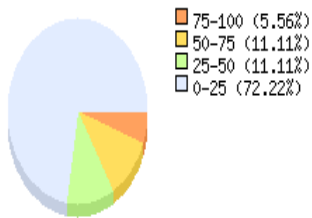
Avg Load (15, 5, 1m):
 34%, 25%, 22%

Localtime:
 2013-12-04 11:02

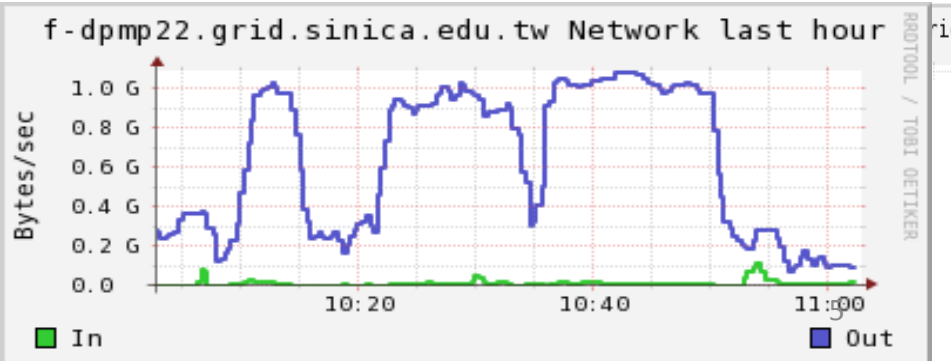
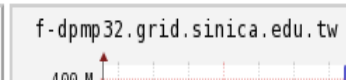
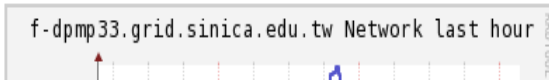


6Gb/s aggregated network throughput

Cluster Load Percentages

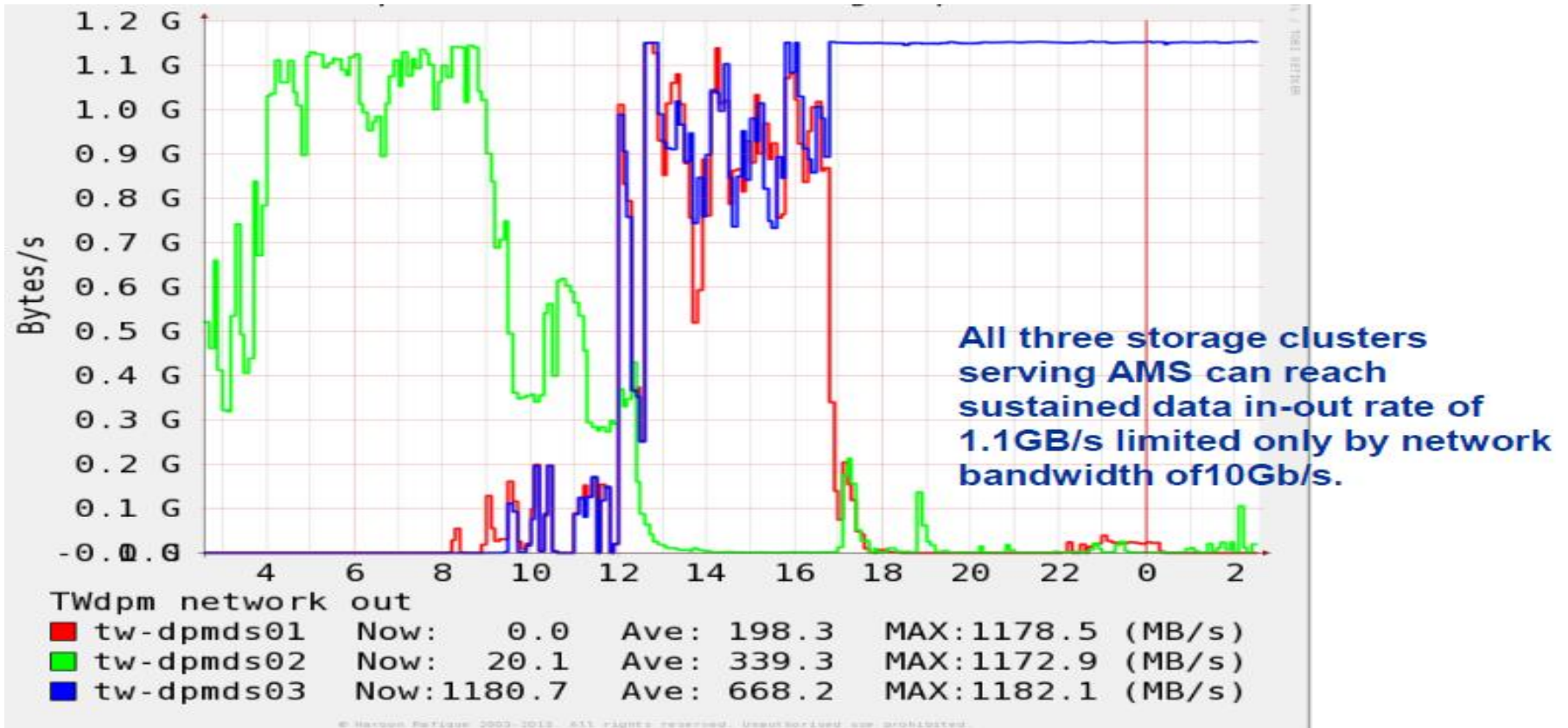


Show Hosts: yes no Taiwan T1T2 - DPM network_report last hour sorted descending Columns 4 Size medium



1GB/s network throughput on one of disk server

Performance (II)



Issues

- Rucio renaming performance
 - ATLAS started re-naming, it was $\sim 0.14\text{Hz}$
 - improved it a lots after switch keepalive on in httpd.conf
- Xrootd redirector crash frequently:
 - happened after upgraded to DPM 1.8.7/xrootd3.3.3
 - daemon was still in running but not handling any requests.
 - we saw “ XrdScheduler: Thread limit has been reached!” in log file(/var/log/xrootd/redir/fedredir_atlas/redir/xrootd.log)
 - tried to increase number of user's processes in Linux not help.
 - David Smith gave us a workaround yesterday, we need to find a chance to restart xrootd and dpm services.
- 140 TB data lost incident on Oct. 28 (will talk more about it later)

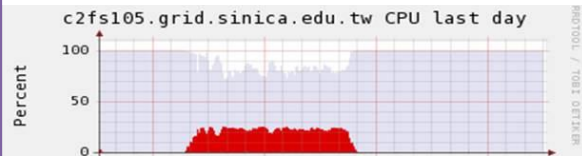
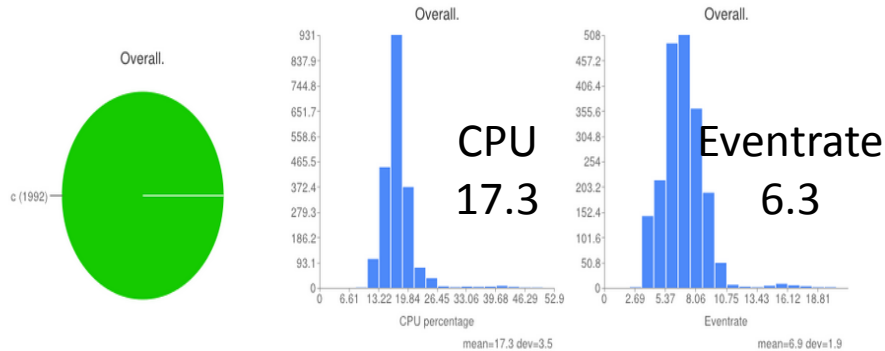
HammerCloud Test

- DPM Version: dpm-xrootd-3.3.5-1.el5.centos, lcgdm-dav-server-0.14.0-1.20131025.1220.el5
- DPM Head node *1
- DPM Disk server *1
 - CPU: Intel(R) Xeon(R) CPU L5520 @ 2.27GHz
 - MEM: 24GB
 - Network: 10GbE
- gLite Work nodes
 - 276 CPU cores (58 blade servers)

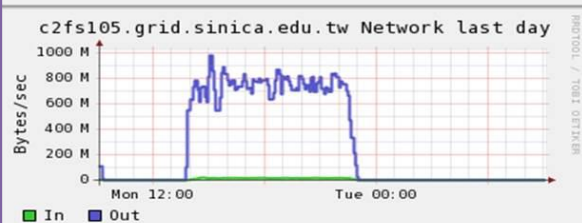
Staging Mode (I)

state	id	host	clouds	start time (CET)	end time (CET)	total jobs
completed	20027905	voatlas65	TW_PANDA	4/11/2013 14:30	4/11/2013 23:30	2406

Input type: PANDA
 Output DS: user.gangarbt.hc20027905.*
 Input DS Patterns:
 mc12_8TeV.165874.Herwigpp_UEEE3_CTEQ6L1_pMSSM_2091483_100FT_neut.merge.AOD.e1469_a159_a171_r3549
 Ganga Job Template: 17.2.7/1727_Panda.tpl
 Athena User Area: 17.2.7/UserAnalysis_v1727.tar.gz
 Athena Option file: 17.2.7/AnalysisSkeleton_topOptions.py
 Template: UA 17.2.7 Panda (default Panda job splitting, special DS)
 View Test Directory (for debugging)

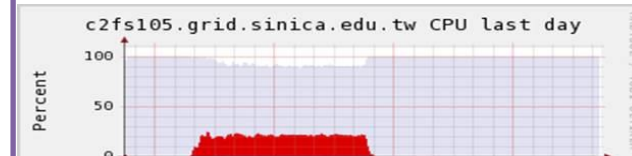
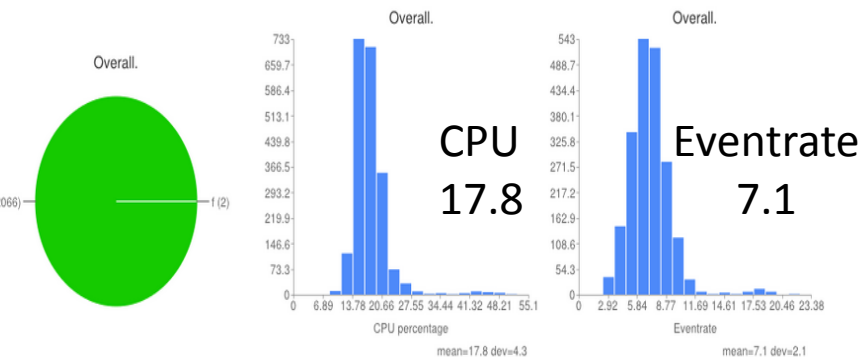


WebDAV

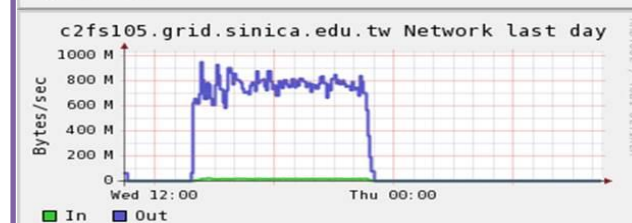


state	id	host	clouds	start time (CET)	end time (CET)	total jobs
completed	20027987	voatlas65	TW_PANDA	6/11/2013 14:30	6/11/2013 23:30	2458

Input type: PANDA
 Output DS: user.gangarbt.hc20027987.*
 Input DS Patterns:
 mc12_8TeV.165874.Herwigpp_UEEE3_CTEQ6L1_pMSSM_2091483_100FT_neut.merge.AOD.e1469_a159_a171_r3549*
 Ganga Job Template: 17.2.7/1727_Panda.tpl
 Athena User Area: 17.2.7/UserAnalysis_v1727.tar.gz
 Athena Option file: 17.2.7/AnalysisSkeleton_topOptions.py
 Template: UA 17.2.7 Panda (default Panda job splitting, special DS)
 View Test Directory (for debugging)



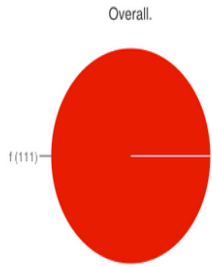
Xrootd



Direct Access (I)

state	id	host	clouds	start time (CET)	end time (CET)	total jobs
completed	20029168	voatlas65	TW_PANDA	29/11/2013 12:00	30/11/2013 3:00	121

Input type: PANDA
 Output DS:
 Input DS Patterns:
 mc12_8TeV.165874.Herwigpp_UEEE3_CTEQ6L1_pMSSM_2091483_100FT_neut.merge.AOD.e1469_a159_a171_r3549*
 Ganga Job Template: 17.2.7/1727_Panda.tpl
 Athena User Area: 17.2.7/UserAnalysis_v1727.tar.gz
 Athena Option file: 17.2.7/AnalysisSkeleton_topOptions.py
 Template: UA 17.2.7 Panda (default Panda job splitting, special DS)
 View Test Directory (for debugging)

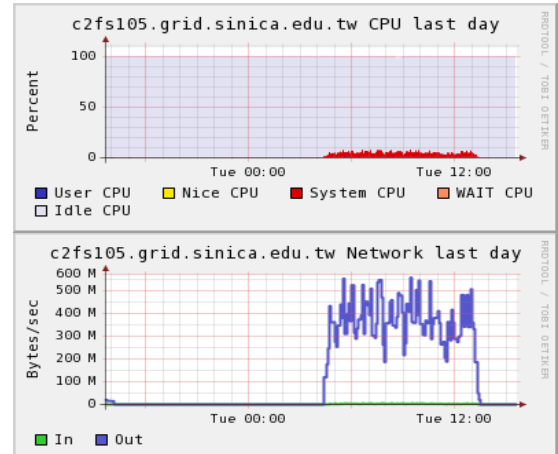
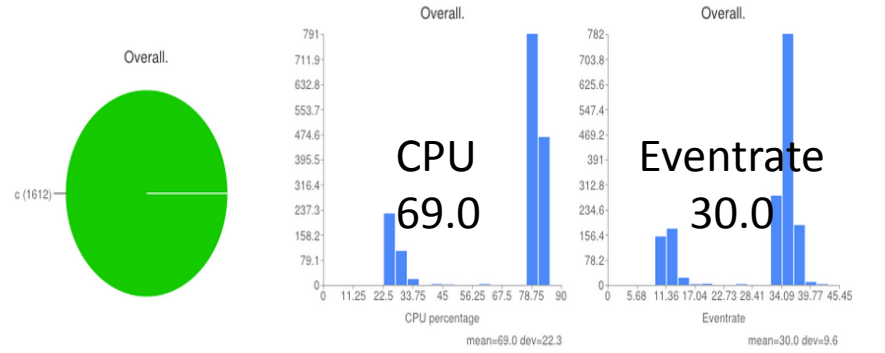


more plots >

WebDAV

state	id	host	clouds	start time (CET)	end time (CET)	total jobs
completed	20029047	voatlas65	TW_PANDA	26/11/2013 5:00	26/11/2013 14:00	1812

Input type: PANDA
 Output DS: user.gangarbt.hc20029047.*
 Input DS Patterns:
 mc12_8TeV.165874.Herwigpp_UEEE3_CTEQ6L1_pMSSM_2091483_100FT_neut.merge.AOD.e1469_a159_a171_r3549*
 Ganga Job Template: 17.2.7/1727_Panda.tpl
 Athena User Area: 17.2.7/UserAnalysis_v1727.tar.gz
 Athena Option file: 17.2.7/AnalysisSkeleton_topOptions.py
 Template: UA 17.2.7 Panda (default Panda job splitting, special DS)
 View Test Directory (for debugging)

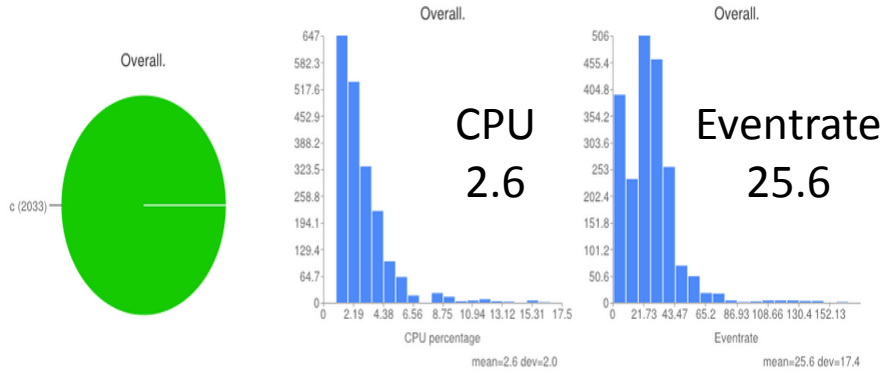


Xrootd

Staging Mode (II)

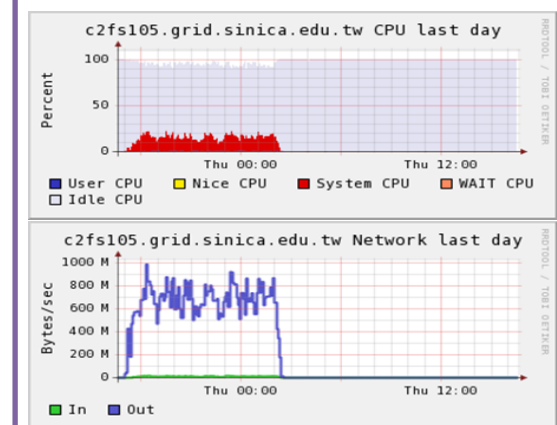
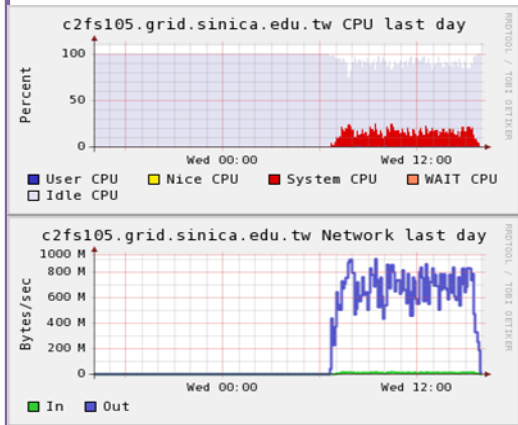
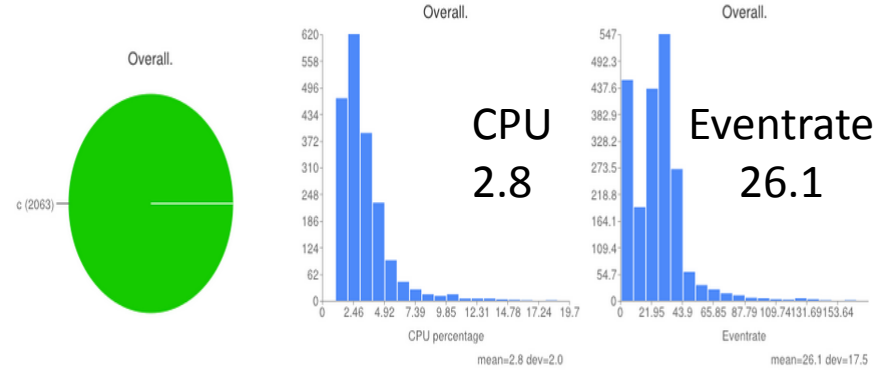
state	id	host	clouds	start time (CET)	end time (CET)	total jobs
completed	20029575	voatlas49	TW_PANDA	11/12/2013 7:30	11/12/2013 16:30	2415

Input type: PANDA
 Output DS: user.gangarbt.hc20029575.*
 Input DS Patterns:
 data12_8TeV.00206497.physics_Muons.merge.NTUP_SMWZ.f453_m1175_p1067_p1141_tid00945658_00
 Ganga Job Template: HWWNtupleCode-00-02-07/HWWNtupleCode-00-02-07.17.2.7-autocache-Panda.tpl
 Athena User Area: HWWNtupleCode-00-02-07/HWWNtupleCode-00-02-07-SL6-autocache.tar.gz
 Athena Option file: HWWNtupleCode-00-02-07/run_MC12a_DATA.sh
 Template: ROOT HWWNtupleCode-00-02-07 DATA p1067 17.2.7 Panda SL6 autocache Panda special dataset
 View Test Directory (for debugging)



state	id	host	clouds	start time (CET)	end time (CET)	total jobs
completed	20029598	voatlas49	TW_PANDA	11/12/2013 18:00	12/12/2013 3:00	2415

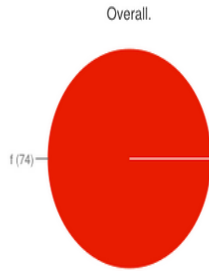
Input type: PANDA
 Output DS: user.gangarbt.hc20029598.*
 Input DS Patterns:
 data12_8TeV.00206497.physics_Muons.merge.NTUP_SMWZ.f453_m1175_p1067_p1141_tid00945658_00
 Ganga Job Template: HWWNtupleCode-00-02-07/HWWNtupleCode-00-02-07.17.2.7-autocache-Panda.tpl
 Athena User Area: HWWNtupleCode-00-02-07/HWWNtupleCode-00-02-07-SL6-autocache.tar.gz
 Athena Option file: HWWNtupleCode-00-02-07/run_MC12a_DATA.sh
 Template: ROOT HWWNtupleCode-00-02-07 DATA p1067 17.2.7 Panda SL6 autocache Panda special dataset
 View Test Directory (for debugging)



Direct Access (II)

state	id	host	clouds	start time (CET)	end time (CET)	total jobs
completed	20029332	voatlas49	TW_PANDA	4/12/2013 5:00	4/12/2013 7:28	74

Input type: PANDA
 Output DS:
 Input DS Patterns:
 data12_8TeV.00206497.physics_Muons.merge.NTUP_SMWZ.f453_m1175_p1067_p1141_tid00945658_00
 Ganga Job Template: HWWNtupleCode-00-02-07/HWWNtupleCode-00-02-07.17.2.7-autocache-Panda.tpl
 Athena User Area: HWWNtupleCode-00-02-07/HWWNtupleCode-00-02-07-SL6-autocache.tar.gz
 Athena Option file: HWWNtupleCode-00-02-07/run_MC12a_DATA.sh
 Template: ROOT HWWNtupleCode-00-02-07 DATA p1067 17.2.7 Panda SL6 autocache Panda special dataset
 View Test Directory (for debugging)

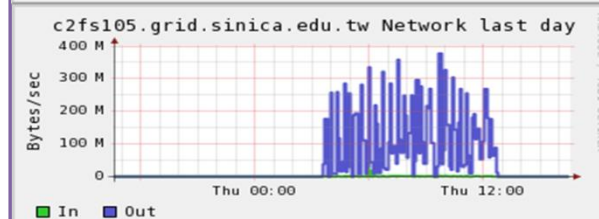
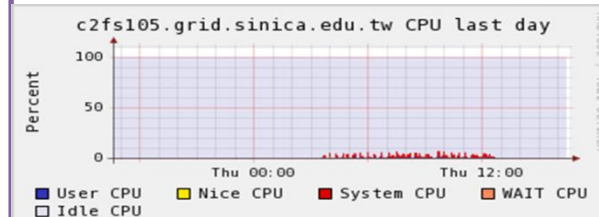
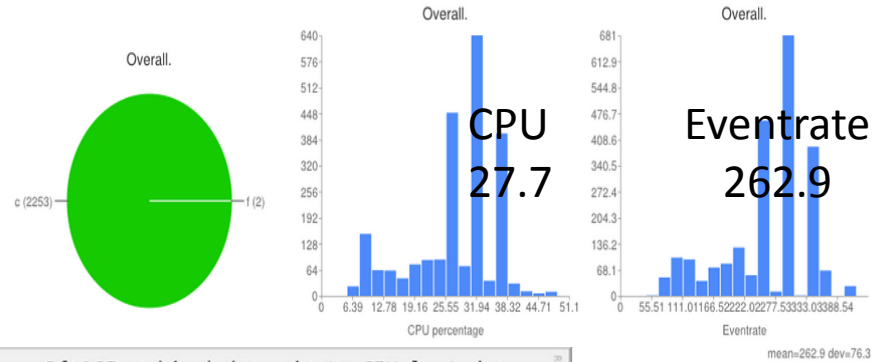


more plots >

WebDAV

state	id	host	clouds	start time (CET)	end time (CET)	total jobs
completed	20029369	voatlas49	TW_PANDA	5/12/2013 4:30	5/12/2013 13:30	2374

Input type: PANDA
 Output DS: user.gangarbt.hc20029369.*
 Input DS Patterns:
 data12_8TeV.00206497.physics_Muons.merge.NTUP_SMWZ.f453_m1175_p1067_p1141_tid00945658_00
 Ganga Job Template: HWWNtupleCode-00-02-07/HWWNtupleCode-00-02-07.17.2.7-autocache-Panda.tpl
 Athena User Area: HWWNtupleCode-00-02-07/HWWNtupleCode-00-02-07-SL6-autocache.tar.gz
 Athena Option file: HWWNtupleCode-00-02-07/run_MC12a_DATA.sh
 Template: ROOT HWWNtupleCode-00-02-07 DATA p1067 17.2.7 Panda SL6 autocache Panda special dataset
 View Test Directory (for debugging)



Xrootd

Summary - HammerCloud

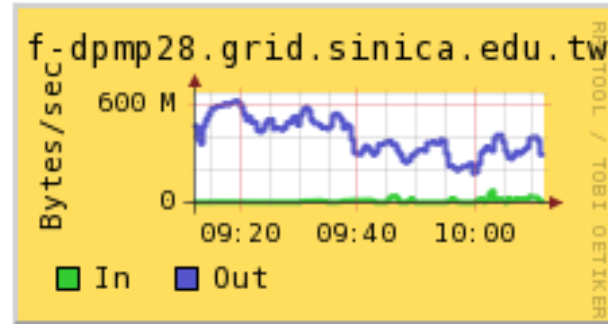
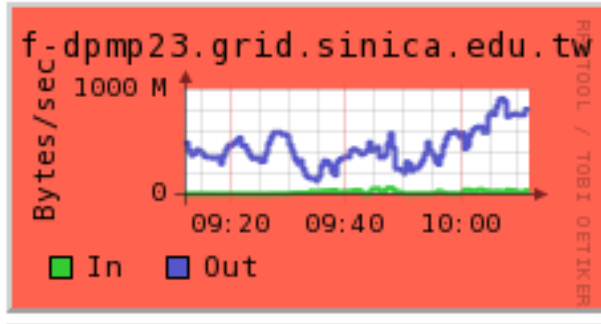
- Direct Access is what we want as we do not have to worry the free local space when dealing with large input and output file.
- More CPU percentage and Eventrate in direct access mode.
- Direct access via http not working in our test (under investigating)

Data loss incident report

ASGC

Story behind the incident

- Observed that some of our DPM disk servers performed poorly.



- They share the same H/W configuration, and were almost all online at the same time.
- We tried to understand and improve the condition.
 - confirmed that our DC network, server H/W, DPM are fine...
- So, we invited our storage vendor to investigate this issue together.

I/O meter benchmark

- Vendor insist on using I/O meter to identify their storage performance as they only trust I/O meter result.
- So, we test I/O meter on our testbed
- Everything was fine on testbed, no any error occurred.
- We then put one of disk servers into RDONLY mode, and ran I/O meter on it.

I/O meter crashed the super blocks??

- Target disk server had 8 storage volumes, 140TB disk space, 1007444 files in total.
- To drain all files from that disk server before doing I/O meter was ever considered.
 - But, we didn't have enough space....
 - We very regret this..
- The benchmark went smoothly at beginning...
- The I/O error happened just right after the I/O meter benchmark was done.
 - All storage volumes went frozen, there were a lot of xfs core dumps happening.
 - We then started the recovery immediately.
- But after we remounted the filesystem successfully, we noticed all filesystems were completely messed....

Recovery

- Do `xfs_repair` to repair all filesystems.
 - But during the repairing, it kept showing inodes were invalid..
 - Then we've got 8 fresh filesystems..
 - Then we panicked...
- Unmount all volumes to avoid any change happened to those filesystems.
- Prepare another clean server and attached volumes to it.
- Use `dd` to create image from original volumes, then:
 - Trying everything we can do to get files back!!

Try recovery by file rescue tools

- **Xfr**
 - Always failed in compiling.
- **Foremost**
 - Can't recognize/recover root file
- **Photorec**
 - Like foremost, it can't recognize/recover root file
- **Scalpel**
 - Rescue file by header and footer.
 - We can customize the file header and footer which we want to recover
 - But it seems that root file only has header.
 - Still, it doesn't work to recover root file.
- **xfs_irecover**
 - Can recover xfs inodes and then get file back.
 - But, it only recovered a few inodes in our case.

Survey other recovery solutions

- A famous studio in south of Taiwan
 - Charge \$\$ per MegaByte....
 - Offer physical, and very raw level scan to hard drives.
 - RAID recovery will be tricky and will charge more.
 - No gaurantee, but since they only charge us by file size, so, we don't recovery any file, they won't charge any fee.
 - The raw level scan is some kind of violent access to device, so, it will damage storage whatever the files can get recovered or not
- Commercial software
 - No gaurantee to recover root files..

Recovery by xfs_irecover

- Confirm that only xfs_irecover could recover files although it couldn't recover all.
- Start to scan all storage volumes
 - Compare checksum between recovered file and DPM DB to identify physical file name.
 - Recovered 9011 files, but only 2829 files can get matched checksum from DPM DB.
- Recovery rate was $\sim 0.2\%$ (2829/1007444)
 - The total recovered size was $\sim 105\text{GB}$, so, it's $\sim 0.07\%$ recovery rate if count it by file size(105GB/140TB)

Analysis

- xfs_irecovery could recover data, which means if we keep super block complete, we should be able to recover more data.
 - We didn't back up xfs journal regularly, we will have regular xfs journal backup from now on.
- Why I/O meter will cause super block being damaged?
 - No answer so far, our vendor - who claim they always use I/O meter to do benchmark also have no clue about this..In fact, they are still trying to avoid discussion about this...
 - They still owe us a explanations...
 - After some arguments. They agree to offer us a testbed in order to do more test .
 - To identify this problem and so do performance issue

Time line of incident

- **15:12 28th Oct, 2013 UTC.** Doing I/O benchmark and got it finished. It's only ran on one disk server . The process was enduring for 10 mins.
- **15:38 28th Oct, 2013 UTC.** We got alarm regarding read/write error to that disk server. Confirm that the disk directories couldn't be accessible. We immediately disabled all partitions on this disk (f-dpmp28.grid.sinica.edu.tw) and started the recovery process.
- **29th Oct, 2013 UTC.** Deploy another server and attached storage to that server, use dd to make image from original partitions in order to try some data rescue tools without bothering original source. But, after several trials, only xfs_irecover could work.
- **1st Nov, 2013 UTC.** Enable FAX to make Atlas job to be able to use xrootd redirector in order to reduce job failure rate. Confirm that the tool was functional and started to scan all partitions. Use checksum to identify the real filename between DPM database and recovered files.
- **5th Nov, 2013 UTC.** The recovery process had been finished. Recovery rate was only 0.07% , had confirmed 140TB data loss. List²³ the lost files to DDM OPS and ATLAS users.

Summary - Data loss incident

- It's quite a hard work to recover 1007444 files...
- Big thanks and also apologies to Atlas DDM ops and all ADC experts!
- To have regular xfs journal backup.
- File replication on different host to protect data.
- Enabled FAX failover in Taiwan did help during data recovery.

TAIWAN-LCG2 data loss - recovery status

- ~1M files lost

Slide from ATLAS ADC on Nov. 27

- 800k on DATADISK

- 350k data files

- declared and almost processed

- 65k recovered (+3k now being transferred)

- 145k lost

- 30k not in LFC/DQ2

- not declared:

- ~100k not yet declared by mistake (under investigation)

- reported today, **shiffters should ignore “source file doesn't exist” for TAIWAN-LCG2 endpoints this week**

- 450k logs

- only important ones declared and processed

- will be declared last